Journal Pre-proof

Development and Validation of a Novel Computed-Tomography Enterography Radiomic Approach for Characterization of Intestinal Fibrosis in Crohn's Disease

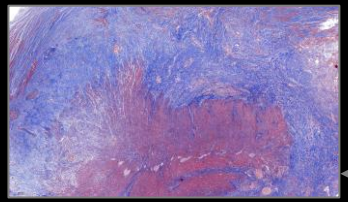
Xuehua Li, MD, Dong Liang, MD, Jixin Meng, MD, Jie Zhou, MD, Zhao Chen, MD, Siyun Huang, MD, Baolan Lu, MD, Yun Qiu, MD, Mark E. Baker, MD, Ziyin Ye, MD, Qinghua Cao, MD, Mingyu Wang, MD, Chenglang Yuan, MD, Zhihui Chen, MD, Shengyu Feng, MD, Yuxuan Zhang, MD, Marietta Iacucci, MD, Subrata Ghosh, MD, Florian Rieder, MD, Canhui Sun, MD, Minhu Chen, MD, Ziping Li, MD, Ren Mao, MD, Bingsheng Huang, MD, Shi-Ting Feng, MD

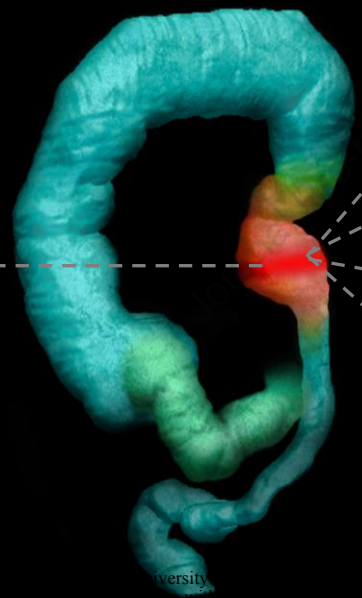
PII: S0016-5085(21)00415-7
DOI: https://doi.org/10.1053/j.gastro.2021.02.027
Reference: YGAST 64146

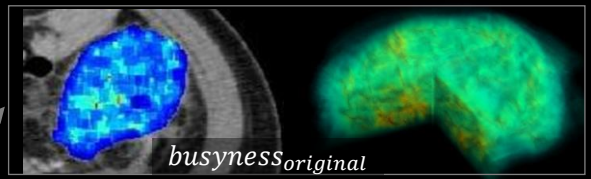
To appear in: *Gastroenterology*
Accepted Date: 9 February 2021

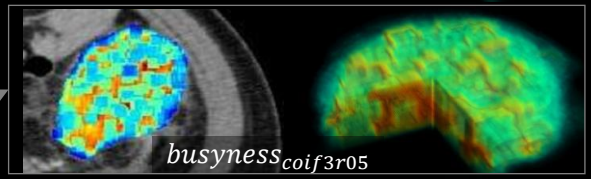
Please cite this article as: Li X, Liang D, Meng J, Zhou J, Chen Z, Huang S, Lu B, Qiu Y, Baker ME, Ye Z, Cao Q, Wang M, Yuan C, Chen Z, Feng S, Zhang Y, Iacucci M, Ghosh S, Rieder F, Sun C, Chen M, Li Z, Mao R, Huang B, Feng S-T, Development and Validation of a Novel Computed-Tomography Enterography Radiomic Approach for Characterization of Intestinal Fibrosis in Crohn's Disease, *Gastroenterology* (2021), doi: https://doi.org/10.1053/j.gastro.2021.02.027.

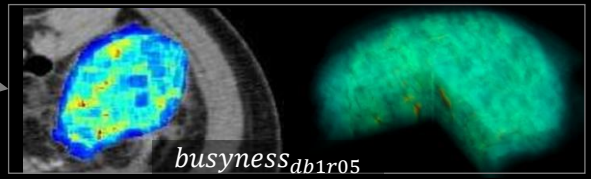
Intestinal fibrosis prediction by CTE-based radiomic model with AUC of 0.83

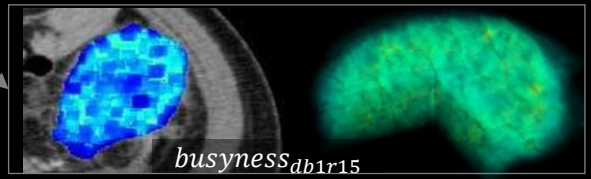












Gastroenterology

Title page

Development and Validation of a Novel Computed-Tomography Enterography Radiomic Approach for Characterization of Intestinal Fibrosis in Crohn's Disease

Short Title: Characterizing bowel fibrosis using radiomics

Authors:

Xuehua Li[1]*, MD, Dong Liang[2]*, MD, Jixin Meng[1]*, MD, Jie Zhou[3], MD, Zhao Chen[4], MD, Siyun Huang[1], MD, Baolan Lu[1], MD, Yun Qiu[5], MD, Mark E. Baker[6], MD, Ziyin Ye[7], MD, Qinghua Cao[7], MD, Mingyu Wang[2], MD, Chenglang Yuan[2], MD, Zhihui Chen[8], MD, Shengyu Feng[2], MD, Yuxuan Zhang[2], MD, Marietta Iacucci[9], MD, Subrata Ghosh[9], MD, Florian Rieder[10], MD, Canhui Sun[1], MD, Minhu Chen[5], MD, Ziping Li[1], MD, Ren Mao[5,10], MD, Bingsheng Huang[2], MD, Shi-Ting Feng[1], MD

*These three authors contributed equally to this work.

Institutions:

[1] Department of Radiology, The First Affiliated Hospital, Sun Yat-Sen University, 58 Zhongshan II Road, Guangzhou 510080, People's Republic of China

[2] Medical AI Lab, School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen 518000, People's Republic of China

[3] Department of Radiology, The Sixth Affiliated Hospital, Sun Yat-Sen University, Yuancun Er Heng Road, NO.26, Guangzhou 510655, People's Republic of China

[4] Department of Medical Imaging Center, Nan Fang Hospital, Southern Medical University, 1838 Guangzhou Avenue North, Guangzhou 510515, People's Republic of China

[5] Department of Gastroenterology, The First Affiliated Hospital, Sun Yat-Sen University, 58 Zhongshan II Road, Guangzhou 510080, People's Republic of China

[6] Section of Abdominal Imaging, Imaging Institute, Digestive Disease Institute and Cancer Institute, Cleveland Clinic, Cleveland, Ohio, USA

[7] Department of Pathology, The First Affiliated Hospital, Sun Yat-Sen University, 58 Zhongshan II Road, Guangzhou 510080, People's Republic of China

[8] Department of Gastrointestinal and Pancreatic Surgery, The First Affiliated Hospital, Sun Yat-Sen University, Guangzhou, People's Republic of China

[9] NIHR Biomedical Research Institute, Institute of Translational Medicine, University of Birmingham and University Hospitals Birmingham NHS Foundation Trust, United Kingdom

[10] Department of Gastroenterology, Hepatology and Nutrition, Digestive Diseases and Surgery Institute, Cleveland Clinic, Cleveland, Ohio, USA

Correspondence:

Ren Mao: Department of Gastroenterology, The First Affiliated Hospital, Sun Yat-Sen University, 58 Zhongshan Road 2nd, Guangzhou 510080, People's Republic of China. Department of Gastroenterology, Hepatology and Nutrition, Digestive Diseases and Surgery Institute, Cleveland Clinic, Cleveland, Ohio, USA. Email: maor5@mail.sysu.edu.cn. Tel: 86-20-87755766-8471, Fax: 86-20-87615805.

Bingsheng Huang: Medical AI Lab, School of Biomedical Engineering, Health Science Center, Shenzhen University. Block A2, Xili Campus of Shenzhen University, 1066 Xueyuan Avenue, Shenzhen 518000, People's Republic of China. Email: huangb@szu.edu.cn. Tel: 86-755-86172208, Fax: 86-755-86171820.

Shi-Ting Feng: Department of Radiology, The First Affiliated Hospital, Sun Yat-Sen University, 58 Zhongshan Road 2nd, Guangzhou 510080, People's Republic of China. Email: fengsht@mail.sysu.edu.cn. Tel: 86-20-87755766-8471, Fax: 86-20-87615805.

Grant Support:

This study was supported by National Natural Science Foundation of China (82070680, 82072002, 81600508, 81970483, 81770654, 81971684, 81771908, 81870451, 81700482), Guangdong Basic and Applied Basic Research Foundation (2020A1515010571), Shenzhen-Hong Kong Institute of Brain Science-Shenzhen Fundamental Research Institutions (2019SHIBS0003) and Nature Science Foundation of Shenzhen (JCYJ20200109114014533). The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Abbreviations:

AUC: Area under the ROC curve

CD: Crohn's disease

CI: Confidence interval

CTDI: CT dose index

CTE: Computed tomography enterography

ICC: Intraclass correlation coefficients

LASSO: Least absolute shrinkage and selection operator

LOOCV: leave-one-out cross-validation

MRI: Magnetic resonance imaging

MRTA: MRI texture analysis

MSE: Mean square errors

NGTDM: Neighbourhood Grey Tone Difference Matrix

RM: Radiomic model

ROC: Receiver operating characteristic

VOI: Volume-of-interest

Disclosures:

All authors declare no competing interests.

Writing Assistance:

None

Author Contributions:

Study conception: XHL, RM, BSH and STF. Data collection: XHL, DL, JXM, JZ, ZC, SYH, BLL, YQ, ZYY, QHC, ZHC, CHS, SYF, YXZ. Data analysis: XHL, DL, JXM, MYW, CLY. Manuscript drafting: XHL and DL. Manuscript editing: RM, BSH, STF, MHC, ZPL, MEB, MI, SG, and FR. All authors reviewed and commented on the manuscript and approved the final version.

Acknowledgments

The authors thank Li Huang, Jinjiang Lin, Yingkui Zhong, Jinfang Du, Zhuangnian Fang, Li Shi and Mengchen Zhang, the radiologists or radiographers from The First Affiliated Hospital of Sun Yat-Sen University, for assisting the data collection. We also thank Zhiyang Zhou (Professor of Radiology, The Sixth Affiliated Hospital of Sun Yat-Sen University) and Yikai Xu (Professor of Radiology, Nanfang Hospital of Southern Medical University) for commenting on this manuscript.

ABSTRACT

Background & Aims: No reliable method for evaluating intestinal fibrosis in Crohn's disease (CD) exists; therefore, we developed a computed-tomography enterography (CTE)–based radiomic model (RM) for characterising intestinal fibrosis in CD.

Methods: This retrospective-multicentre study included 167 CD patients with 212 bowel lesions (training, 98 lesions; test, 114 lesions) who underwent preoperative CTE and bowel resection at one of the three tertiary referral centres from January 2014 through June 2020. Bowel fibrosis was histologically classified as none-mild or moderate-severe. In the training cohort, 1454 radiomic features were extracted from venous-phase CTE, and a machine learning–based RM was developed based on the reproducible features using logistic regression. The RM was validated in an independent external test cohort recruited from three centres. The diagnostic performance of RM was compared with two radiologists' visual interpretation of CTE using receiver operating characteristic (ROC) curve analysis.

Results: In the training cohort, the area under the ROC curve (AUC) of RM for distinguishing moderate-severe from none-mild intestinal fibrosis was 0.888 (95% confidence interval [CI]: 0.818–0.957). In the test cohort, the RM showed robust performance across three centres with an AUC of 0.816 (95% CI: 0.706–0.926), 0.724 (95% CI: 0.526–0.923), and 0.750 (95% CI: 0.560–0.940), respectively. Moreover, the RM was more accurate than visual interpretations by either radiologist (#1 AUC=0.554; #2 AUC=0.598; both $P<0.001$) in the test cohort. Decision curve analysis showed that the RM provided a better net benefit to predicting intestinal fibrosis than the radiologists.

Conclusion: A CTE-based RM allows for accurate characterisation of intestinal fibrosis in CD.

Keywords: Crohn's disease; Fibrosis; Radiomics; Computed tomography enterography

INTRODUCTION

Fibrostenosis is a serious complication of Crohn's disease (CD) affecting approximately half of all patients.[1, 2] Excessive collagen deposition within the bowel wall is the main feature of CD fibrosis. Fibrosis and inflammation typically coexist in the affected intestine, and current anti-inflammatory therapies used in CD cannot effectively prevent or reverse established intestinal fibrosis.[3] CD fibrostenosis is typically treated with surgical resection. Hence, accurate characterisation of intestinal fibrostenosis is crucial for managing patients with CD.[1] As transmural fibrosis is not discernible by endoscopic visualisation or partial-thickness biopsy, cross-sectional imaging has been explored as a non-invasive alternative for characterising fibrostenosis.[4] Emerging imaging techniques such as magnetisation transfer magnetic resonance imaging (MRI)[5-7] and ultrasonic elastography[8] are reportedly able to accurately assess intestinal fibrosis. However, the interobserver variation and subjectivity of magnetisation transfer MRI and the operator level dependence of ultrasonic elastography are of concern when using these techniques in clinical practice.[9] To date, there is no consensus on the most reliable method for evaluating intestinal fibrosis in CD.

Computed tomography enterography (CTE) is a widely used imaging modality for CD; however, it lacks accuracy in detecting intestinal fibrosis.[1, 10, 11] The conventional visual interpretation on CTE images may not be sufficient for full and objective evaluation of lesion features, which may underestimate CTE's ability to assess intestinal fibrosis. Recently, radiomics, which involves computer-based extraction of large amounts of high-dimensional features from medical images, has been explored to overcome these limitations.[12, 13] It can uncover disease characteristics that may not be detected by visual inspection of images. Radiomics has been used in colorectal cancer studies[14-18] but could potentially be applied to other digestive diseases including CD. Few studies have used radiomics to assess CD.[19, 20] Tabari *et al.*[19] recently reported that texture analysis of contrast-enhanced MRI can detect bowel fibrosis in CD. Although this study involved inherent limitations (e.g. it only included 25 paediatric patients and lacked independent external validation), it showed the feasibility of using

radiomics analysis of cross-sectional imaging for assessing intestinal fibrosis. However, feature extraction from MRI is more complex and less reproducible than that from CT images, indicating that the results from Tabari *et al.* may not be easily reproducible on another cohort.[21] To our knowledge, no study has investigated the efficacy of radiomic CT analysis for evaluating intestinal fibrosis in adults with CD.

This multicentre study developed and externally validated a CTE-based radiomic model (RM) to characterise intestinal fibrosis in CD and compared its capability with radiologist-performed visual interpretation of CTE images. We hypothesised that different levels of collagen deposition in fibrostenosis lead to a distinct contrast-enhanced pattern on the bowel wall on CTE in ways that might be detectable by radiomics but may not be visually apparent; thus, radiomics may provide better intestinal fibrosis diagnostic accuracy.

MATERIALS AND METHODS

Patients and study design

This retrospective multicentre study, which included 167 consecutive eligible patients with CD from three tertiary referral centres in China from January 2014 through June 2020, was approved by the institutional ethics review board of The First Affiliated Hospital of Sun Yat-Sen University, which waived the requirement for informed consent. The training cohort was recruited from The First Affiliated Hospital of Sun Yat-Sen University (Centre 1) while the external test cohort involved patients from Centre 1, The Sixth Affiliated Hospital of Sun Yat-Sen University (Centre 2), and Nanfang Hospital of Southern Medical University (Centre 3).

The inclusion criteria were as follows: *(a)* available preoperative abdominal and pelvic CTE within 3 months of surgery for bowel strictures or fistula/abscess; and *(b)* availability of a histopathologic surgical specimen of a segment of intestine corresponding to a matching abnormality on CTE. The exclusion criteria were as follows: *(a)* inadequate imaging quality resulted from breathing artifact; *(b)* targeted bowel segment located at an anastomosis because the natural imaging features of anastomosis would be changed by the medical intervention; *(c)* not readily identifiable intestinal contour on CTE due to severe perienteric effusion, intestinal adhesion or intestinal peristalsis; or *(d)* emergency surgery without enhanced-CT scanning (Fig. 1).

The final training cohort included 87 patients (62 men, 25 women; mean age: 34.36±11.89 years) with 98 bowel segments. The final test cohort included 80 patients (55 men, 25 women; mean age: 34.19±11.41 years) with 114 bowel segments. The demographic and clinical characteristics of the training and test cohort are shown in Table 1.

Reference standard for intestinal fibrosis

Matched radiological and surgical evaluation was performed as previously reported[5, 22] by a radiologist (C.S.) and a surgeon (Z.C.). Briefly, a detailed description of the features of each lesion

(e.g. location, length, and stricture) were reviewed in surgical and pathological records to ensure that the same lesions were analysed in cases of patients with multifocal intestinal involvement,[22] or the anatomic location of target areas was documented with respect to defined anatomic landmarks (e.g. ileocecal valve or appendix) in the operating room.[5] The detailed information is shown in the Supplementary materials. The histologic sections from three centres were assessed using the same criteria by a pathologist (Q.C., with 10 years of experience in bowel pathology) who was blinded to clinical and radiological information. The pathologist graded the bowel fibrosis on Masson's trichrome section and bowel inflammation on haematoxylin and eosin section from 0 to 4 using a semi-quantitative scoring system (Supplementary Table 1).[10] Histologic fibrosis and inflammation scores were categorised in none-mild (scores, 0–2) or moderate-severe (scores, 3–4) groups.[23] The histologic fibrosis and inflammation scores of both cohorts are shown in Supplementary Table 2.

CTE examination and radiologist-performed visual interpretation of intestinal fibrosis on CTE images

All patients underwent preoperative CTE using one of the five CT scanners (Supplementary materials) in three centres. On plain and dual-phase enhancement CTE, two radiologists (X.L. and S.H., with 10 and 8 years of experience in abdominal CT, respectively) unaware of pathological and radiomic data separately assessed intestinal fibrosis based on the following three traits[3, 4, 11, 24]: lumen narrowing; pre-stricture dilation; iso-density or mildly progressive pattern after enhancement (imaging-feature definitions in Supplementary Table 3). The observation result of each imaging feature was recorded as a score, of which '0' refers to absence and '1' to presence (Supplementary Table 4). The kappa statistic was calculated to determine interobserver agreement between the two radiologists. The diagnostic performance of radiologist-performed visual interpretation was evaluated using logistic regression, and the three imaging features were used as the input of the regression model.

The logistic regression models established with radiologist-performed visual interpretation are shown in the Supplementary materials.

Volume-of-interest (VOI) segmentation and radiomic feature extraction and selection

Similar to other bowel radiomics studies,[14, 19] we selected contrast-enhanced venous-phase bowel images for radiomics analysis. Three-dimensional VOIs of the bowel lesions in the training and test cohorts were manually segmented on the venous-phase CT images by a radiologist (J.M., with 4 years of experience in abdominal CT), using an open-source medical imaging software (MITK, version 2018.04; https://www.mitk.org/). The VOIs were drawn along the lesion contour on each transverse section until the full lesion was captured excluding the intestinal lumen. The completed VOIs were used as a mask to select the voxels that encompassed the lesion.

To investigate the inter- and intra-observer reproducibility of extracting radiomic features, two radiologists (B.L. with 7 years of experience in abdominal CT, and J.M.) undertook additional VOI segmentation that consisted of 61 bowel segments in the test cohort recruited from Centre 1 using the same tool and environment settings. The time interval between two readings of one radiologist (J.M.) was over 6 months. The average time to manually segment one VOI by the radiologists was 4.73±3.40 minutes. Intraclass correlation coefficients (ICCs) were calculated from a two-way random effects model to determine the inter- and intra-observer reliability. Only radiomic features that had excellent reliability (ICC>0.90) were considered robust.[25]

Radiomic features extraction was performed using in-house software developed in MATLAB (version R2018a; https://www.mathworks.org/). Within each VOI, three wavelet transforms (Daubechies, Coiflets, and Symlets) with three different ratios (0.5, 1.5, and 2.0) between lowpass (approximation) and high-pass (detail) components were performed on the original image (10 images in total). Next, 24 shape features that assessed the spatial properties of three-dimensional VOI were automatically extracted from original images. Then, 37 first-order statistical features and 106 texture

features were calculated on the 10 images, resulting in (37+106)×10=1430 features. In total, 1454 radiomic features were available for each VOI (Supplementary materials).

To reduce the dimensionality and select the best subset of features, the following feature selection strategies were used: *(a)* Selecting features that were under a significance level of 0.1 in differentiating none-mild and moderate-severe intestinal fibrosis in the training cohort according to univariate analysis; *(b)* The least absolute shrinkage and selection operator (LASSO) of logistic regression model,[26] with penalty tuning conducted by 10-fold cross-validation, was applied to select fibrosis-related features (Supplementary materials).

To investigate the utility of radiomic features for characterizing the spatial variation of fibrosis within the bowel strictures, feature maps of selected features were generated using voxel-based radiomics analysis. The VOI for each subject was used to limit the calculation of local feature values. Radiomic features were calculated in smaller blocks with a kernel radius of three pixels throughout the VOI, and the values were assigned to the centre pixel.

Radiomic model development in the training cohort and validation in the test cohort

A binary classification RM model was built for distinguishing between none-mild and moderate-severe intestinal fibrosis based on the selected radiomic features by using logistic regression. Logistic regression is a machine learning algorithm aiming to find the best fit of a logistic function that is diagnostically reasonable to describe the relationship between target and predictor variables. To avoid overfitting and select the best parameters for the diagnostic model, we conducted leave-one-out cross-validation (LOOCV) in the training cohort. In the training procedure of this LOOCV, a grid search method was used to determine the optimal parameters made of *C* (inverse of regularisation strength), *max_iter* (maximum number of iterations taken for the solvers to converge), and *tol* setting (tolerance for stopping criteria) in the logistic regression model.[27] The overall predicting performance

in the training cohort was assessed based on all probabilities of testing samples in each fold during cross-validation.

The validation of the RM was performed using the same metrics in the completely independent test cohort recruited from Centres 1–3. The logistic regression model, built on the training cohort using the optimal parameters and selected features, was then applied to the test cohort.

The performance of the diagnostic model was evaluated using the area under the receiver operating characteristic (ROC) curve (AUC) and by performing calibration using the Hosmer–Lemeshow goodness-of-fit test. Model development was performed in the Python environment (version 3.6; https://www.python. org/) using the Scikit-learn package (version 0.22; https://www.scikit-learn.org/). Our radiomics analysis workflow is presented in Fig. 2 and detailed in the Supplementary materials.

Stratification analysis of radiomic model's diagnostic performance

Given that inflammation and fibrosis often coexist in the intestinal strictures in CD,[5] evaluation of RM's ability to assess intestinal fibrosis in different degrees of inflammation is necessary. To address this issue, the diagnostic performance of the RM was compared between the bowel walls with moderate-severe inflammation and those with none-mild inflammation. Additionally, studies have suggested that small bowel CD and colonic CD represent two distinct disease entities.[28-30] Therefore, the difference in the diagnostic performance of the RM between small bowel strictures and colonic strictures is worthy of exploration. We also investigated the differences in the diagnostic efficiency of the RM between different CT scanners and between bowel strictures with and without penetrating diseases, to assess its stability and robustness. The penetrating diseases were evaluated based on radiologist-performed visual interpretation[4] and intraoperative findings.

Clinical utility of radiomic model and radiologist-performed visual interpretation

Decision curve analysis was used to measure the clinical utility of the RM and radiologist-performed visual interpretation; a decision analytic measure called net benefit of the model was calculated for each possible threshold probability by summing the benefits (proportion of true-positive) and subtracting the harms (proportion of false-positive), weighting the latter by the relative harm of a false-positive and a false-negative result.[31] The net benefit values of the diagnostic models were standardised for prevalence.

Statistical analysis

Sample size

A sample size of at least 26 bowel lesions (13 none-mild and 13 moderate-severe fibrosis) was required in the training and test cohorts based on the following input and assumption: power, 80%; two-sided significance level, 0.05; alternative hypothesis of the AUC, 0.800, compared with the null hypothesis of the AUC, 0.500, and an allocation ratio of sample sizes in the negative and positive groups of 1.[32] Therefore, sample sizes of 98 (33 none-mild and 65 moderate-severe fibrosis) in the training cohort and 114 (32 none-mild and 82 moderate-severe fibrosis) in the test cohort were sufficient to detect an AUC different from 0.500 with 80% power if the true AUC was above 0.800.

Diagnostic performance

Normally distributed data are expressed as mean ± standard deviation, and non-normally distributed data are presented as median (interquartile range [IQR]). The differences in radiomic features, clinical characteristics, and histologic scores between none-mild and moderate-severe intestinal fibrosis were calculated with Student's *t*-test, Welch's *t*-test, or Mann–Whitney *U*-test according to the data distribution. A bivariate correlation analysis between radiologist-performed visual interpretation and histologic fibrosis scores was performed using Kendall's τ_b correlation. The discrimination

performances of the RM and radiologist-performed visual interpretation were evaluated by the ROC curve and AUC value, reported with corresponding 95% confidence interval (CI). An AUC of 0.50–0.69 was considered mild, 0.70–0.89 moderate, 0.90–1.00 good. The accuracy, sensitivity, and specificity were calculated from the ROC curve according to the cut-off value that maximises the Youden index, equal to sensitivity + specificity - 1. Paired and unpaired ROC curves between different models were compared using the DeLong method.[33, 34] The calibration performances of the models were assessed using the Hosmer–Lemeshow goodness-of-fit test. Statistical analysis was performed in Python and R (version 3.6.3; *https://www.r-project.org/*) environments. A two-sided $P<0.05$ was considered significant, except for the univariate analysis ($P<0.1$).

RESULTS

Radiomic feature selection and model development

Radiomic feature selection

Of the 1454 radiomic features extracted from the training cohort, 127 that were significantly different between none-mild and moderate-severe intestinal fibrosis were considered for the following analysis. Subsequently, the optimised α parameter (the weight of L1-term in LASSO, and determined by 10-fold cross-validation) was used to build the LASSO model and four fibrosis-related features were selected based on the training cohort (Supplementary Fig. 1A–B). The four selected features were all *Busyness* based on *Neighbourhood Grey Tone Difference Matrix* (*NGTDM*) on four different images (original image and three wavelet-transformed images). Details on the selected radiomic features are provided in Supplementary Table 5. The heatmap of selected features in both cohorts is shown in Supplementary Fig. 1C–D.

Inter- and intra-observer reproducibility of extracting radiomics features

Among the 1454 features, 60 with zero variance were excluded. The median ICCs of the other 1394 radiomic features for inter- and intra-observer agreement assessment were 0.985 (IQR, 0.965–0.993) and 0.988 (IQR, 0.968–0.996) respectively, suggesting marked reliability of the radiomic features. The four selected features in our RM also showed robust inter- and intra-observer reproducibility, with ICCs ranging from 0.977 to 0.998 (all P<0.001) and from 0.963 to 0.997 (all P<0.001), respectively.

Radiomic model development

A logistic regression model was built in the training cohort and then applied to the test cohort. The outcome of the RM was generated as a function of a linear combination of the four selected features weighted by their respective coefficients using the logistic function. The outcome is a value between zero and one, denoting the predicted probability of the input bowel segment. The formula is:

$$predicted\ probability = \frac{1}{1 + e^{-g(x)}},$$

$$g(x) = \left(-4.2393 \times busyness_{original} + 0.4272 \times busyness_{coif3r05} + 6.1727 \times busyness_{db1r05} - 0.2293 \times busyness_{db1r15} + 0.0019\right) \times 10^{-4}$$

where $busyness_{original}$ is the *Busyness* value of the original image, $busyness_{coif3r05}$ is the *Busyness* value of the image transformed with the Coiflets wavelet and a compression ratio of 0.5, $busyness_{db1r05}$ and $busyness_{db1r15}$ are the values of *Busyness* on the Daubechies-transformed image with compression ratios of 0.5 and 1.5, respectively. The *Busyness* measures the spatial frequency of intensity changes from one pixel to its neighbours. A higher value of *Busyness* indicates a 'busy' image, with rapid changes in intensity between pixels and its neighbourhood.[35] Fig. 3A and Fig. 3B show images of representative patients with moderate-severe and none-mild bowel fibrosis, respectively, on CTE and relevant *Busyness* feature maps overlaid on venous-phase images.

Diagnostic performance of the radiomic model in the training and test cohorts

The predicted probabilities of our RM for each bowel lesion in both cohorts are provided in Fig. 4A–B. With LOOCV, the RM predicted intestinal fibrosis with an AUC of 0.888 (P<0.001) in the training cohort (Table 2).

In the multicentre validation, our RM exhibited a satisfied performance for predicting intestinal fibrosis both in the combined test cohort (AUC=0.832; P<0.001; Table 2) and in each single centre (AUC=0.816, 0.724, 0.750, respectively; Fig. 4C; Table 3). There were no significant differences in the AUC among the three centres (all P>0.400).

No significant difference in the AUC was observed between the training cohort and the total test cohort according to DeLong's test (P=0.331), suggesting considerable efficacy and robustness of our RM (Fig. 4D). The Hosmer–Lemeshow test yielded a χ^2 value of 8.997 (P=0.343) and 11.006 (P=0.201) in training cohort and total test cohort, suggesting good RM fit (Fig. 4E).

Comparison of diagnostic performance of the radiomic model and radiologist-performed visual interpretation

There were positive correlations between lumen narrowing and histologic fibrosis scores in the test cohort (radiologist 1: Kendall's τ_b=0.226, *P*=0.015; radiologist 2: Kendall's τ_b=0.250, *P*=0.007) but not in the training cohort (radiologist 1: Kendall's τ_b=0.052, *P*=0.606; radiologist 2: Kendall's τ_b=-0.044, *P*=0.664). No significant correlation was found between histologic fibrosis scores and either pre-stricture dilation or enhancement pattern in the training and test cohorts (Kendall's τ_b=-0.148-0.163; all *P*>0.05).

As shown in Table 2, two radiologists' visual interpretation showed poor diagnostic performance for differentiating varying degrees of intestinal fibrosis (radiologist 1: AUC=0.508; *P*=0.892; radiologist 2: AUC=0.567; *P*=0.281) in the training cohort. The RM significantly outperformed visual interpretation by both radiologist 1 (AUC=0.888 vs. 0.508, *P*<0.001) and radiologist 2 (AUC=0.888 vs. 0.567, *P*<0.001) in diagnosis of intestinal fibrosis (Fig. 4F). The decision curve (Fig. 4G) showed that, if the threshold probability is approximately over 20%, the RM had a higher net benefit than the radiologists' visual interpretation in the training cohort, indicating better performance of the RM in terms of clinical application.

In the test cohort, the diagnostic performance of the RM in differentiating moderate-severe from none-mild intestinal fibrosis was also significantly higher than that of the radiologists' visual interpretation (radiologist 1: AUC=0.554; *P*=0.354; radiologist 2: AUC=0.598; *P*=0.071) (both *P*<0.005 for DeLong's test; Fig. 4H). The decision curve analysis (Fig. 4I) in the test cohort showed that the RM always provided a better net benefit to predict intestinal fibrosis than the radiologists' visual interpretation.

Diagnostic performances of radiomic model in different inflammatory severity, different CD locations, different CT scanners, and bowel strictures with and without penetrating diseases

In different inflammatory severity--In the training cohort, the bowel lesions were categorised as none-mild inflammatory segments (16.33%, 16/98) or moderate-severe inflammatory segments (83.67%, 82/98) based on their histologic inflammation scores (Supplementary Table 2). As shown in Table 2, the RM achieved a moderate diagnostic efficacy for differentiating moderate-severe from none-mild fibrosis in bowel walls with none-mild inflammation (AUC=0.836; *P*=0.010) or those with moderate-severe inflammation (AUC=0.890; *P*<0.001). No significant difference in the RM's AUC for diagnosing fibrosis was found between these subgroups (*P*=0.702).

In the test cohort, 43 bowel segments were in none-mild inflammation (37.72%, 43/114) and 71 in moderate-severe inflammation (62.28%, 71/114) (Supplementary Table 2). The RM had moderate diagnostic efficacy for differentiating varying degrees of fibrosis in bowel walls with none-mild inflammation (AUC=0.829; *P*<0.001) or those with moderate-severe inflammation (AUC=0.785; *P*<0.001) (Table 2). No significant difference in the RM's AUC for diagnosing fibrosis was found between these subgroups (*P*=0.662).

In different CD locations--No significant differences in the AUC of the RM for differentiating varying degrees of fibrosis were found between the small bowel strictures and colonic strictures in either the training or test cohort (*P*=0.569, 0.810, respectively) (Supplementary Fig.2, Supplementary Table 6).

In different CT scanners--No significant differences in the AUC of the RM for differentiating varying degrees of fibrosis were shown between different CT scanners in either cohort (all *P*>0.150) (Supplementary Fig.3, Supplementary Table 7).

In bowel strictures with and without penetrating diseases—No significant differences in the AUC of the RM for differentiating varying degrees of fibrosis were found between bowel segments with and

without penetrating diseases in either the training or test cohort (*P*=0.744, 0.149, respectively) (Supplementary Fig.4, Supplementary Table 8).

DISCUSSION

We developed a novel CTE-based RM for assessment of intestinal fibrostenosis in CD. Our findings indicate that an acceptably high diagnostic classification efficiency between none-mild and moderate-severe intestinal fibrosis in patients with CD can be achieved using our RM, which showed remarkable robustness in agreement among three independent institutions and in bowel segments with different inflammatory severity. Additionally, our RM significantly outperformed radiologist-performed visual interpretation in differentiating varying grades of intestinal fibrosis in CD.

The development of radiomics has altered the way radiologists use medical imaging for diagnosis.[12] Radiomics has been successfully applied to evaluate fibrosis of other organs (e.g. the liver).[36] However, to our knowledge, only one study[19] has used radiomics analysis to evaluate intestinal fibrosis on MRI. CTE, which has high spatial resolution, may contribute to accurate delineation of bowel contour and VOI segmentation. To date, no study has used CTE-based radiomics to evaluate intestinal fibrosis in adult patients with CD.

Our results filled this gap. In our study, the CTE-based RM was able to differentiate moderate-severe from none-mild intestinal fibrosis with moderate diagnostic accuracy. Unlike the study by Tabari *et al.*[19], in which three texture features (i.e. mean, skewness and entropy) were selected to predict stricture fibrosis, our RM consisted of four busyness texture features with excellent interobserver reproducibility. The different imaging modalities (MRI vs. CTE) and predictive tasks (strictures with/without fibrosis vs. strictures with none-mild/moderate-severe fibrosis) in these two studies were the possible reasons leading to the difference of selected features. These busyness features in our study, signifying rapid changes in intensity from one pixel to its neighbour,[35] may reflect the spatial heterogeneity of collagen deposition within the intestinal strictures. For example, the features of $busyness_{coif3r05}$ and $busyness_{db1r05}$, which reflect the busyness on the transformed images with stronger high-frequency components, have the positive correlation with the predicted probability of

the RM, while another two features calculated from the images with original or stronger low-frequency components have negative weights in the formula.

Given that various CT scanners and scanning protocols exist in different institutions, assessing RM robustness is key to enabling its widespread use. Studies have reported the repeatability and reproducibility of radiomic signatures from different CT vendors or different scanning parameters with inconsistent results.[37-39] Here, although the CTE images were obtained from five different CT scanners with respective imaging parameters in three centres, our RM still showed remarkable robustness. In the multicentre validation, the RM accuracy in diagnosing intestinal fibrosis did not significantly vary among the three centres and the diagnostic RM accuracy of each centre nearly matched the overall accuracy. The patients enrolled from these three tertiary referral centres were living in all Chinese regions and the CTE images were acquired over a 6-year period. Hence, our RM had good generalisation.

Of note, the performance of this RM for diagnosing intestinal fibrosis was not affected by the severity of the superimposed inflammation within the same bowel segment. As known, inflammation and fibrosis often coexist in the affected bowel walls in CD,[5] and the coexisted inflammation increases the difficulty in detecting intestinal fibrosis by conventional imaging analytical strategies. Here, the RM accuracy for diagnosing fibrosis in moderate-severe inflammatory segments was similar to that in none-mild inflammatory segments, indicating the stability and reliability of this RM for assessing intestinal fibrosis. Additionally, the diagnostic performance of the RM was not affected by the presence of penetrating diseases and by the locations of CD (small bowel or colon). As the RM showed satisfactory accuracy, excellent robustness, and high stability in diagnosing intestinal fibrosis in the training and test cohorts, it may be valuable of clinical generalisation.

Notably, compared to radiologist-performed visual interpretation, the RM showed a significantly higher accuracy in differentiating varying grades of intestinal fibrosis with a better performance in clinical application. The RM was only based on venous-phase contrast-enhanced CTE, whereas

radiologist-performed visual interpretation used additional evaluation of plain and arterial-phase contrast-enhanced CTE. Radiomics analysis has greatly improved the ability of CTE for assessing intestinal fibrosis, remedying the limitation of the conventional strategy in interpretation of intestinal fibrosis on CT images, as coexisted inflammation can mask transmural fibrosis. Although we used preoperative data to develop the RM, this radiomic signature has potential as a practical imaging biomarker to detect bowel fibrosis during clinical management and to personalise the treatment of bowel strictures for patients with CD. Moreover, it might be a reliable and non-invasive approach to identify the endpoint of clinical trials of anti-fibrosis drugs and to promote the progress of anti-fibrosis drugs for patients with CD, because the lack of accurate diagnostic methods for bowel fibrosis may be slowing this progress.[3, 40] This CTE-based radiomic method has attractive application prospects and can benefit patients with CD in primary health care institutions, because other more advanced imaging equipment (e.g. MRI scanners) is expensive, whereas radiomics analysis continues to decrease in cost.

The study had several limitations. First, the sample sizes in the training and test cohorts were limited. This is generally a feature of studies using surgical histopathology as a reference standard of bowel fibrosis in CD. However, to our knowledge, our study included the largest sample size reported to date in a radiomics analysis of bowel fibrosis in CD. In future prospective study, more centres' data collection could sufficiently increase the sample size to strengthen the conclusions. Second, we used CTE rather than MR enterography to develop the RM. MR enterography is the preferred examination for follow-up of patients with CD as it is non-radiative. However, the complexity of imaging technology and parameters of MR enterography increases the difficulty of developing an MR enterography–based RM with sufficient robustness, especially in the case of limited sample sizes. Although CTE still has the potential risk of radiation, it is recommended as the preferred method for CD patients with acute symptoms or with suspected complex intra-abdominal penetrating diseases.[4] Furthermore, modern low-dose CT scanners have the potential to considerably reduce the ionizing radiation exposure for CD patients.[41] Development of an RM based on MR enterography is worth

further study; however, developing an RM using CTE first, which would be used for making critical decision regarding management and not for repeated use, is a good beginning in the application of artificial intelligence for diagnosing intestinal fibrosis in patients with CD. Additionally, in this retrospective study, few patients underwent both CTE and MR enterography during the same period. Hence, we did not compare the diagnostic performance of a CTE-based RM with that of an MR enterography–based RM. Third, this RM might not be applicable for evaluating the fibrotic severity of bowel strictures with blurring of the contour, because the diagnostic accuracy of the RM may decrease when the bowel contour is difficult to recognize and the outer or lumen of the intestine might be incorrectly drawn into the area of interest. Last, manual VOI segmentation for the RM is time-consuming, and the development of an automated or semi-automated segmentation tool based on deep learning is necessary. Deep learning is an emerging segmentation-free technique in computational radiology, and it outperforms radiomics in a variety of tasks. However, more CTE data are required to build a robust model during deep learning model training.[42] Moreover, due to their end-to-end learning design, deep learning models often appear to be black boxes, absorbing data and generating output without clear explanations behind their prediction.[43] In contrast, the RM in our study could be better understood due to the pre-defined radiomic features and simple regression formula. Since our RM has achieved a considerable diagnostic performance, a deep learning–based segmentation tool, which can encourage the use of the RM in daily clinical practice with optimum reproducibility, is worthy of future exploration.

In conclusion, our RM allows for accurate characterisation of intestinal fibrosis in CD and offers notable advantages over radiologist-performed visual interpretation for differentiating varying grades of intestinal fibrosis.

References

1. Bettenworth D, Bokemeyer A, Baker M, et al. Assessment of Crohn's disease-associated small bowel strictures and fibrosis on cross-sectional imaging: a systematic review. Gut 2019; 68:1115-1126.

2. Rieder F, Latella G, Magro F, et al. European Crohn's and Colitis Organisation topical review on prediction, diagnosis and management of fibrostenosing Crohn's disease. J Crohns Colitis 2016; 10:873-885.

3. Rieder F, Bettenworth D, Ma C, et al. An expert consensus to standardise definitions, diagnosis and treatment targets for anti-fibrotic stricture therapies in Crohn's disease. Aliment Pharmacol Ther 2018; 48:347-357.

4. Bruining DH, Zimmermann EM, Loftus EJ, et al. Consensus recommendations for evaluation, interpretation, and utilization of computed tomography and magnetic resonance enterography in patients with small bowel Crohn's disease. Gastroenterology 2018; 154:1172-1194.

5. **Li XH**, **Mao R**, Huang SY, et al. Characterization of degree of intestinal fibrosis in patients with Crohn disease by using magnetization transfer MR imaging. Radiology 2018; 287:494-503.

6. **Meng J**, **Huang S**, Sun C, et al. Comparison of three magnetization transfer ratio parameters for assessment of intestinal fibrosis in patients with Crohn's disease. Korean J Radiol 2020; 21:290-297.

7. **Fang ZN**, **Li XH**, Lin JJ, et al. Magnetisation transfer imaging adds information to conventional MRIs to differentiate inflammatory from fibrotic components of small intestinal strictures in Crohn's disease. Eur Radiol 2020; 30:1938-1947.

8. **Chen YJ**, **Mao R**, Li XH, et al. Real-time shear wave ultrasound elastography differentiates fibrotic from inflammatory strictures in patients with Crohn's disease. Inflamm Bowel Dis 2018; 24:2183-2190.

9. Lin S, Lin X, Li X, et al. Making qualitative intestinal stricture quantitative: embracing radiomics in IBD. Inflamm Bowel Dis 2020; 26:743-745.

10. Adler J, Punglia DR, Dillman JR, et al. Computed tomography enterography findings correlate with tissue inflammation, not fibrosis in resected small bowel Crohn's disease. Inflamm Bowel Dis 2012; 18:849-856.

11. Chiorean MV, Sandrasegaran K, Saxena R, et al. Correlation of CT enteroclysis with surgical pathology in Crohn's disease. Am J Gastroenterol 2007; 102:2541-2550.

12. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology 2016; 278:563-577.

13. Lambin P, Leijenaar R, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 2017; 14:749-762.

14. Yang L, Dong D, Fang M, et al. Can CT-based radiomics signature predict KRAS/NRAS/BRAF mutations in colorectal cancer? Eur Radiol 2018; 28:2058-2067.

15. Cui Y, Yang X, Shi Z, et al. Radiomics analysis of multiparametric MRI for prediction of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. Eur Radiol 2019; 29:1211-1220.

16. **Liu Z**, **Zhang XY**, **Shi YJ**, et al. Radiomics analysis for evaluation of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. Clin Cancer Res 2017; 23:7253-7262.

17. Shi L, Zhang Y, Nie K, et al. Machine learning for prediction of chemoradiation therapy response in rectal cancer using pre-treatment and mid-radiation multi-parametric MRI. Magn Reson Imaging 2019; 61:33-40.

18. **Horvat N**, **Veeraraghavan H**, Khan M, et al. MR imaging of rectal cancer: radiomics analysis to assess treatment response after neoadjuvant therapy. Radiology 2018; 287:833-843.

19. Tabari A, Kilcoyne A, Jeck WR, et al. Texture analysis of magnetic resonance enterography contrast enhancement can detect fibrosis in Crohn disease strictures. J Pediatr Gastroenterol Nutr 2019; 69:533-538.

20. Makanyanga J, Ganeshan B, Rodriguez-Justo M, et al. MRI texture analysis (MRTA) of T2-weighted images in Crohn's disease may provide information on histological and MRI disease activity in patients undergoing ileal resection. Eur Radiol 2017; 27:589-597.

21. Kumar V, Gu Y, Basu S, et al. Radiomics: the process and the challenges. Magn Reson Imaging 2012; 30:1234-1248.

22. Wagner M, Ko HM, Chatterji M, et al. Magnetic resonance imaging predicts histopathological composition of ileal Crohn's disease. J Crohns Colitis 2018; 12:718-729.

23. **Zhang MC**, **Li XH**, Huang SY, et al. IVIM with fractional perfusion as a novel biomarker for detecting and grading intestinal fibrosis in Crohn's disease. Eur Radiol 2019; 29:3069-3078.

24. Rimola J, Planell N, Rodriguez S, et al. Characterization of inflammation and fibrosis in Crohn's disease lesions by magnetic resonance imaging. Am J Gastroenterol 2015; 110:432-440.

25. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016; 15:155-163.

26. Tibshirani R. Regression shrinkage and selection via the lasso: a retrospective. J R Statist Soc B 2011; 73:273-282.

27. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12: 2825-2830.

28. Newman B, Silverberg MS, Gu X, et al. CARD15 and HLA DRB1 alleles influence susceptibility and disease localization in Crohn's disease. Am J Gastroenterol 2004; 99:306-315.

29. **Cuthbert AP**, **Fisher SA**, Mirza MM, et al. The contribution of NOD2 gene mutations to the risk and site of disease in inflammatory bowel disease. Gastroenterology 2002; 122:867-874.

30. Dulai PS, Singh S, Vande CN, et al. Should we divide Crohn's disease into ileum-dominant and isolated colonic diseases? Clin Gastroenterol Hepatol 2019; 17:2634-2643.

31. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making 2006; 26:565-574.

32. Obuchowski NA. ROC analysis. Am J Roentgenol 2005; 184:364-372.

33. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988; 44:837-845.

34. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 2011; 12:77.

35. Amadasun M, King R. Textural features corresponding to textural properties. IEEE Trans Syst Man Cybern 1989; 19:1264-1274.

36. Park HJ, Lee SS, Park B, et al. Radiomics analysis of gadoxetic acid-enhanced MRI for staging liver fibrosis. Radiology 2019; 290:380-387.

37. Beig N, Khorrami M, Alilou M, et al. Perinodular and intranodular radiomic features on lung CT images distinguish adenocarcinomas from granulomas. Radiology 2019; 290:783-792.

38. Berenguer R, Pastor-Juan M, Canales-Vazquez J, et al. Radiomics of CT features may be nonreproducible and redundant: influence of CT acquisition parameters. Radiology 2018; 288:407-415.

39. Mackin D, Fave X, Zhang L, et al. Measuring computed tomography scanner variability of radiomics features. Invest Radiol 2015; 50:757-765.

40 D'Haens G, Rieder F, Feagan BG, et al. Challenges in the pathophysiology, diagnosis and management of intestinal fibrosis in inflammatory bowel disease. Gastroenterology 2019; S0016-5085(19)41035-4.

41. Rosenfeld G, Brown J, Vos PM, et al. Prospective comparison of standard- versus low-radiation-dose CT enterography for the quantitative assessment of Crohn disease. Am J Roentgenol 2018; 210:W54-W62.

42. Shen D, Wu G, Suk HI. Deep learning in medical image analysis. Annu Rev Biomed Eng 2017; 19:221-248.

43. Wang F, Casalino LP, Khullar D. Deep learning in medicine—promise, progress, and challenges. JAMA Intern Med 2019; 179:293-294.

Author names in bold designate shared co-first authorship.

Figure legends

Fig. 1. Flow diagram of the study population.

(CD, Crohn's disease; Centre 1, The First Affiliated Hospital of Sun Yat-Sen University; Centre 2, The Sixth Affiliated Hospital of Sun Yat-Sen University; Centre 3, Nanfang Hospital of Southern Medical University)

Fig. 2. The radiomics analysis workflow.

(VOI, volume-of-interest; LASSO, least absolute shrinkage and selection operator)

Fig. 3. Images show the radiomics analysis in (**A**) a 17-year-old girl with CD with moderate-severe fibrotic stricture (fibrosis score, 4; inflammation score, 3; predicted probability of RM, over 99.9%), and (**B**) a 34-year-old woman with CD with none-mild fibrotic stricture (fibrosis score, 1; inflammation score, 1; predicted probability of RM, below 0.1%). In each part, image (***a***) shows volume rendering reconstruction (blue-green area: intestinal tract; pink area: colonic stricture; red area: VOI in the stricture); axial contrast-enhanced venous-phase CTE image (***b***) shows a stricture of colon (arrowhead); images show Masson's trichrome (***c***) and haematoxylin and eosin staining (***d***), respectively (×40 magnification). Illustrations of the four *busyness* radiomic features overlaid on CTE images and their three-dimensional stereogram are as follows: (***e***) $busyness_{original}$, (***f***) $busyness_{coif3r05}$, (***g***) $busyness_{db1r05}$, and (***h***) $busyness_{db1r15}$. To generate feature maps, the features were calculated for each voxel in the segmented VOIs with a kernel radius of 3 pixels, and the value was assigned to the centre pixel.

(RM, radiomic model; CD, Crohn's disease, CTE; computed-tomography enterography; VOI, volume-of-interest)

Fig. 4. Diagnostic performance of the radiomic model in the training and test cohorts (**A-E**). Plots show the predicted probabilities of the RM for each bowel lesion in (**A**) the training cohort (black arrow refers to the probability [below 0.1%] of the patient with none-mild fibrotic stricture presented in Fig.3B), and in (**B**) the test cohort (black arrow refers to the probability [over 99.9%] of the patient with moderate-severe fibrotic stricture presented in Fig.3A). Plot (**C**) shows the ROC curves for the RM in three independent centres of the test cohort. Plots show the ROC curves (**D**) and the calibration curves (**E**) for the RM in both cohorts.

Comparison of diagnostic performance between the radiomic model and the radiologist-performed visual interpretation of CTE images (**F-I**). The RM shows a significantly higher AUC in differentiating moderate-severe from none-mild intestinal fibrosis than the AUCs of the two radiologists' visual interpretations (all $P<0.005$ according to DeLong's test) in (**F**) the training cohort and (**H**) the test cohort. The decision curve analysis shows that, (**G**) if the threshold probability is approximately over 20%, the RM has a higher net benefit than the radiologists' visual interpretation of the CTE images in the training cohort, while (**I**) RM always has a higher net benefit than the radiologists' visual interpretation in the test cohort.

(RM, radiomic model; ROC, receiver operating characteristic; CTE; computed-tomography enterography; AUC, area under the receiver operating characteristic curve)

Table 1. Demographic and clinical characteristics in patients with Crohn's disease

	Training cohort (87 patients)	Test cohort				*P*§
		Total (80 patients)	Centre 1* (27 patients)	Centre 2* (32 patients)	Centre 3* (21 patients)	
Sex, n (male/female)	62/25	55/25	19 / 8	23/9	13/8	0.723
Age, years (mean±SD)	34.36±11.89	34.19±11.41	32.89±10.92	32.38±10.18	38.62±13.06	0.926
Disease duration, months (median [IQR])	36 (12-72)	36 (14.25-72)	37 (18-72)	58.50(25.25-105)	18 (6-30)	0.413
Interval between CT and surgery, days (median [IQR])	14 (9-28)	14 (10-24)	19 (11-34)	14 (10.25-21.50)	11 (7.50-15)	0.604
Smoking, n (%)	21 (24.14%)	15 (18.75%)	6 (22.22%)	2 (6.25%)	7 (33.33%)	0.398
BMI (median [IQR])	16.80 (15.10-18.90)	18.20 (16.06-20.47)	18 (16-19.20)	18.60 (17.08-21.18)	18.50 (15.49-20.94)	0.067
Previous surgery history, n (%)						0.654
CD-associated bowel resection	12 (13.79%)	10 (12.50%)	0	6 (18.75%)	4 (19.05%)	--
Perianal surgery	12 (13.79%)	9 (11.25%)	5 (18.52%)	4 (12.50%)	0	--
Type of surgery in the present study, n (%)						0.001
Ileocolic resection	42 (48.28%)	18 (22.5%)	13 (48.15%)	1 (3.13%)	4 (19.05%)	--
Partial small bowel resection	19 (21.84%)	13 (16.25%)	3 (11.11%)	4 (12.50%)	6 (28.57%)	--
Partial colon resection	9 (10.34%)	13 (16.25%)	7 (25.93%)	4 (12.50%)	2 (9.52%)	--
Ileocolic + Partial small bowel resection	12 (13.79%)	14 (17.50%)	2 (7.41%)	12 (37.50%)	0	--
Ileocolic + Partial colon resection	3 (3.45%)	12 (15%)	0	5 (15.63%)	7 (33.33%)	--
Partial small bowel + Partial colon resection	2 (2.30%)	10 (12.50%)	2 (7.41%)	6 (18.75%)	2 (9.52%)	--
Location of Specimen, n (%)	98 #	114 #	61 #	32 #	21 #	0.331

Small bowel	92 (93.88%)	90 (78.95%)	56 (91.81%)	21 (65.62%)	13 (61.90%)	--
Colon	6 (6.12%)	24 (21.05%)	5 (8.20%)	11 (34.38%)	8 (38.10%)	--
CDAI, n (%)						＜0.001
< 150 (remission)	13 (14.94%)	21 (26.25%)	4 (14.81%)	9 (33.33%)	8 (38.10%)	--
150-220 (mild activity)	13 (14.94%)	29 (36.25%)	9 (33.33%)	14 (43.75%)	6 (28.57%)	--
220-450 (moderate activity)	54 (62.07%)	28 (35%)	13 (48.15%)	8 (25%)	7 (33.33%)	--
> 450 (severe activity)	7 (8.05%)	2 (2.50%)	1 (3.70%)	1 (3.70%)	0	--
CRP, mg/L (median [IQR])	26.10 (8.31-77.60)	16.29 (5.63-54.77)	29.50 (9.42-56.48)	8.88 (3.01-18.05)	49.52 (8.94-95.96)	0.814
ESR, mm/h (median [IQR])	37 (19-53)	35 (13.50-50)	40 (13-51)	22 (14-45)	42 (11.70-80)	0.610

BMI, Body mass index; CDAI, Crohn's disease activity index; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; IQR, interquartile range; SD, standard deviation

Number of involving bowel segments

∗ Centre 1, The First Affiliated Hospital of Sun Yat-Sen University; Centre 2, The Sixth Affiliated Hospital of Sun Yat-Sen University; Centre 3, Nanfang Hospital of Southern Medical University

§Comparison between training cohort and total test cohort

Table 2. Diagnostic performances of the radiomic model and radiologists' visual interpretation in differentiating moderate-severe from none-mild intestinal fibrosis in the training cohort and the total test cohort

	AUC (95%CI)	Accuracy	Sensitivity	Specificity	*P*
Training cohort (n=98[#])					
Radiomic Model					
Overall	0.888 (0.818-0.957)	0.857	0.815	0.939	<0.001
Inflammatory severity					
None-mild	0.836 (0.580-1.000)	0.813	0.818	0.800	0.010
Moderate-severe	0.890 (0.813-0.966)	0.866	0.815	0.964	<0.001
Visual Interpretation					
Radiologist 1	0.508 (0.388-0.629)	0.582	0.646	0.455	0.892
Radiologist 2	0.567 (0.448-0.686)	0.531	0.477	0.636	0.281
Test cohort (n=114[#])					
Radiomic Model					
Overall	0.832 (0.745-0.919)	0.833	0.902	0.656	<0.001
Inflammatory severity					
None-mild	0.829 (0.683-0.926)	0.721	0.750	0.704	<0.001
Moderate-severe	0.785 (0.671-0.874)	0.901	0.939	0.400	<0.001
Visual Interpretation					
Radiologist 1	0.554 (0.458-0.647)	0.570	0.561	0.594	0.354
Radiologist 2	0.598 (0.502-0.689)	0.649	0.585	0.813	0.071

Note. AUC, area under receiver operating characteristic curve; CI, confidence interval

Accuracy, sensitivity, and specificity of the radiomic model in training and test cohorts were calculated with the cut-off value of 0.811, which maximizes the Youden index in the training cohort.

P value is the significance level of comparison of AUC with that of random case (AUC=0.5)

#Number of involving bowel segments

Table 3. Diagnostic performances of the radiomic model and radiologists' visual interpretation in the independent test cohort of three centres

	AUC (95%CI)	Accuracy	Sensitivity	Specificity	*P*
Centre 1[#] (n=61[*])					
Radiomic Model	0.816 (0.706-0.926)	0.754	0.818	0.679	<0.001
Visual Interpretation					
Radiologist 1	0.558 (0.413-0.703)	0.574	0.515	0.643	0.628
Radiologist 2	0.536 (0.390-0.683)	0.590	0.606	0.571	0.439
Centre 2[#] (n=32[*])					
Radiomic Model	0.724 (0.526-0.923)	0.906	1.000	0	0.027
Visual Interpretation					
Radiologist 1	0.546 (0.361-0.722)	0.719	0.759	0.333	0.834
Radiologist 2	0.511 (0.329-0.691)	0.375	0.345	0.667	0.962
Centre 3[#] (n=21[*])					
Radiomic Model	0.750 (0.560-0.940)	0.857	0.900	0	0.409
Visual Interpretation					
Radiologist 1	0.650 (0.441-0.859)	0.381	0.350	1.000	0.620
Radiologist 2	0.525 (0.194-0.856)	0.333	0.300	1.000	0.934

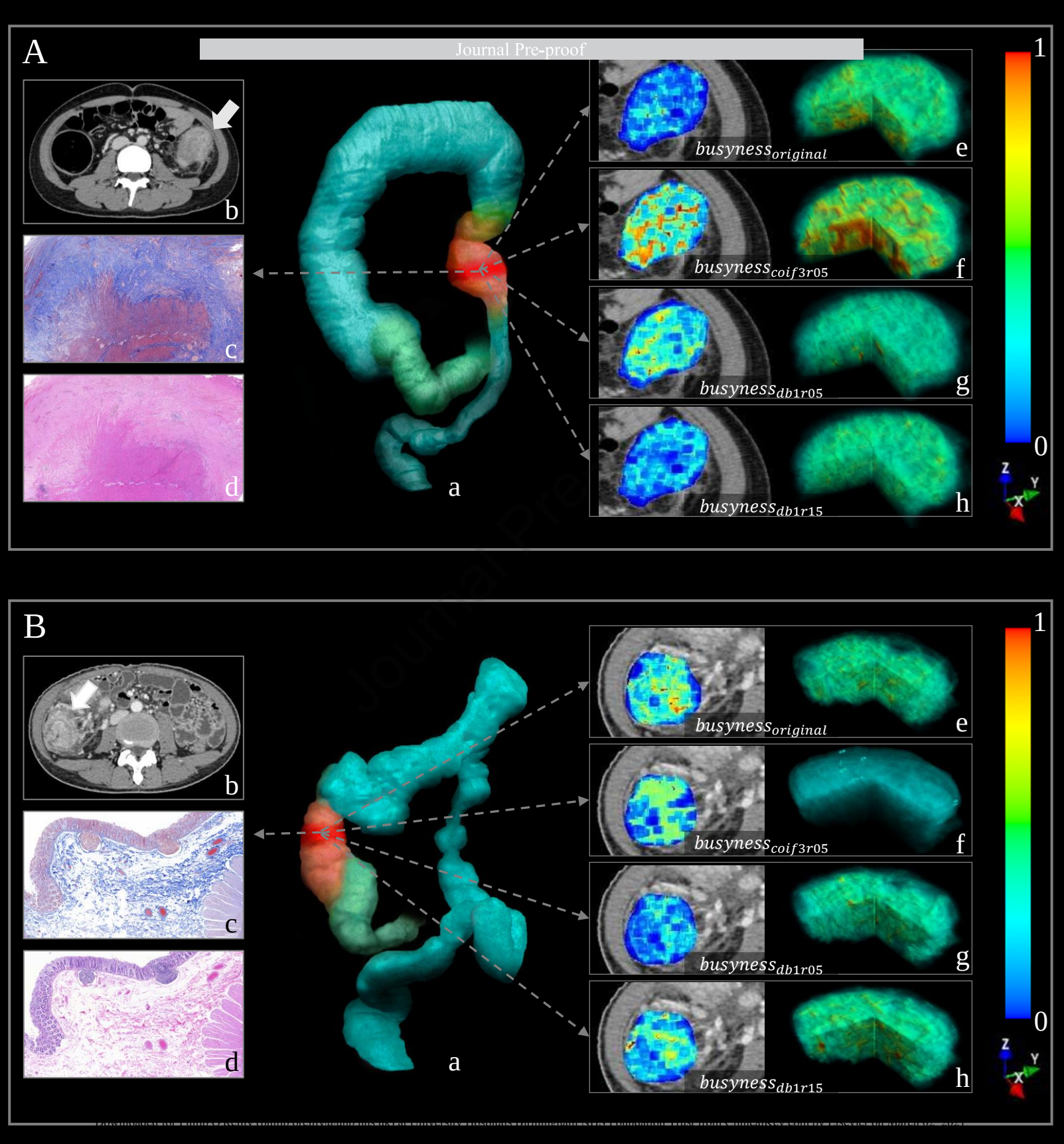
Note. AUC, area under receiver operating characteristic curve; CI, confidence interval

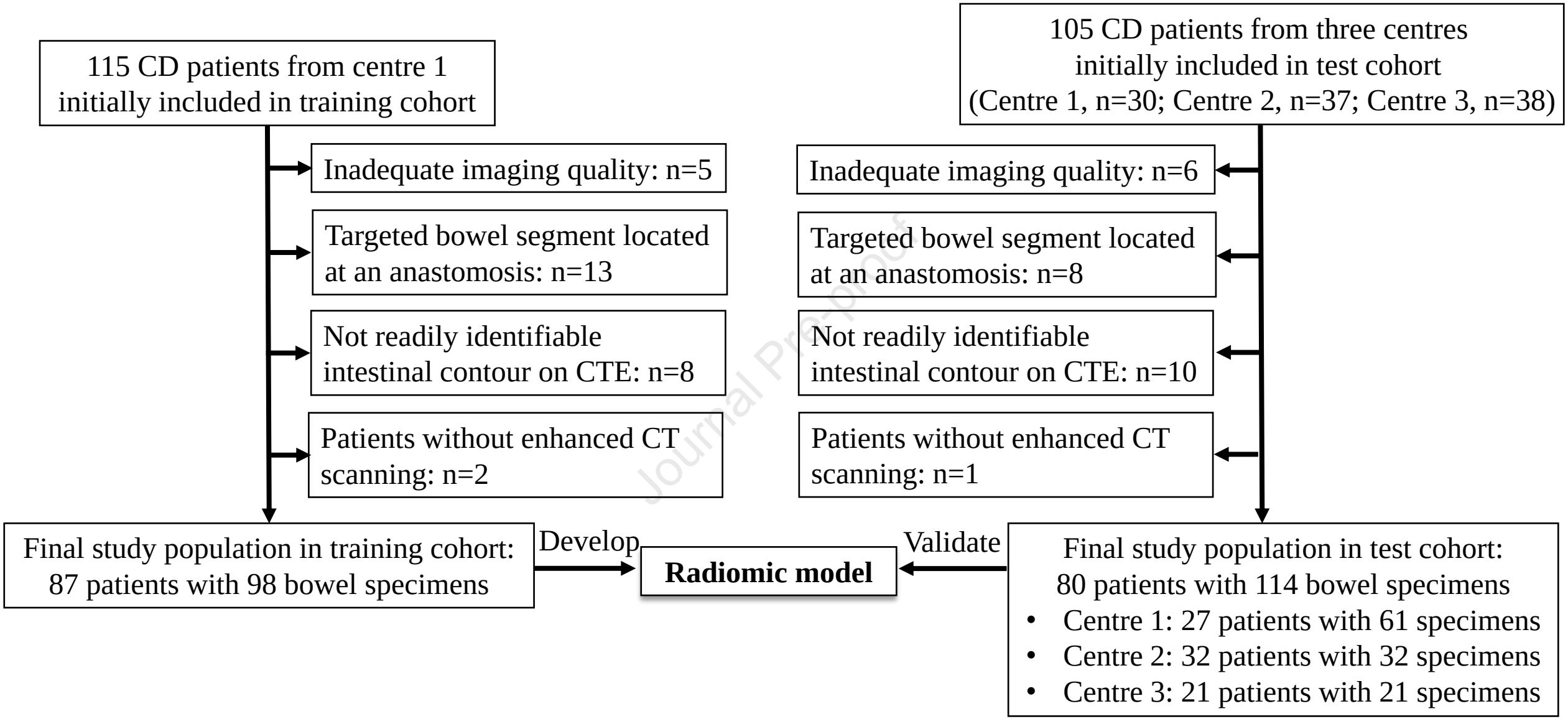
Accuracy, sensitivity, and specificity of the radiomic model in three centres were calculated with the cut-off value of 0.811, which maximizes the Youden index in the training cohort.

P value is the significance level of comparison of AUC with that of random case (AUC=0.5)

* Number of involving bowel segments

Centre 1, The First Affiliated Hospital of Sun Yat-Sen University; Centre 2, The Sixth Affiliated Hospital of Sun Yat-Sen University; Centre 3, Nanfang Hospital of Southern Medical University

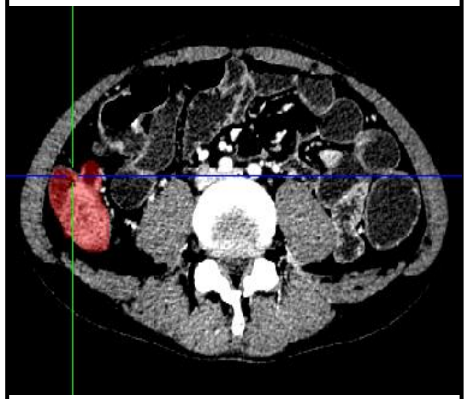




CT Images

Histologic Scoring

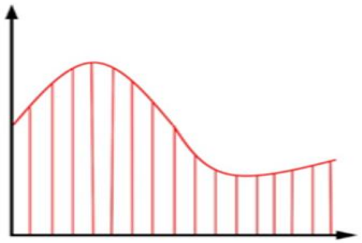
VOI Segmentation



Feature Extraction

Volume/Shape Feature

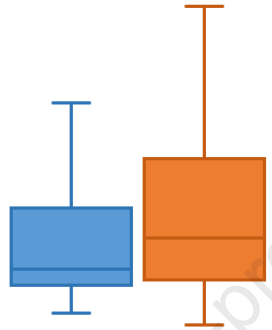
First-order Feature



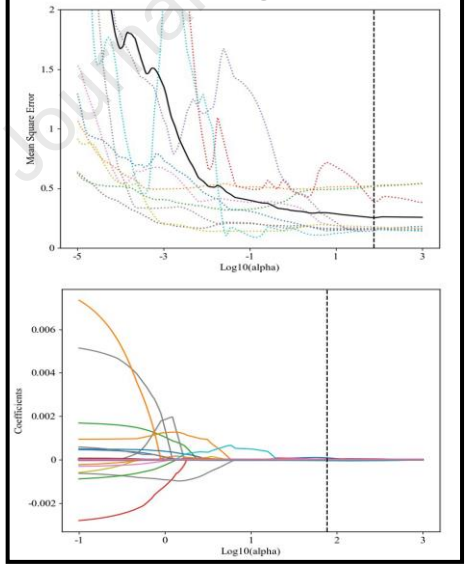
Texture Feature

Feature Selection

Univariate Analysis

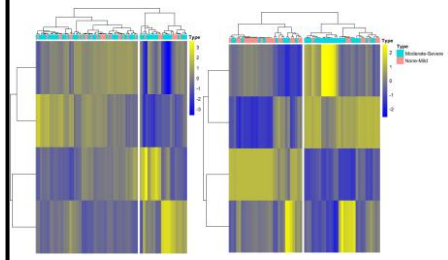


LASSO Regression



Model Development

Selected Features



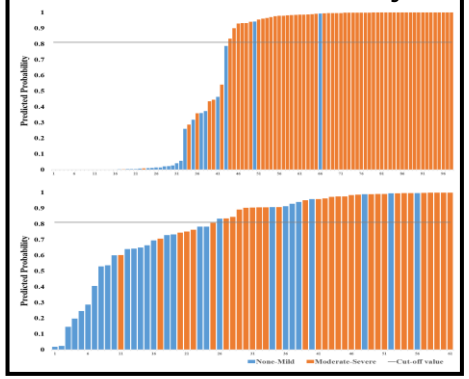
Logistic Regression

$$predicted\ probability = \frac{1}{1+e^{-g(x)}},$$

$$where\ g(x) = \left(-4.2393 \times busyness_{original} + 0.4272 \times busyness_{coif3r05} + 6.1727 \times busyness_{db1r05} - 0.2293 \times busyness_{db1r15} + 0.0019\right) \times 10^{-4}$$

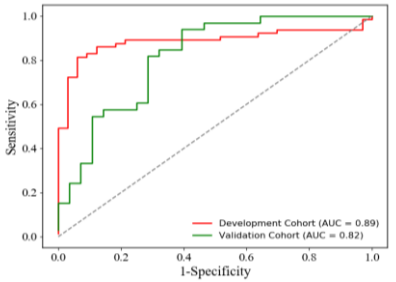
Note.
$busyness_{original}$ is the *Busyness* value of the original image,
$busyness_{coif3r05}$ is the *Busyness* value of the image transformed with Coiflets wavelet and the compression ratio of 0.5,
$busyness_{db1r05}$ is the *Busyness* value of the Daubechies-transformed image with a compression ratio of 0.5,
and $busyness_{db1r15}$ is the *Busyness* value of the Daubechies-transformed image with a compression ratio of 1.5.

Predicted Probability

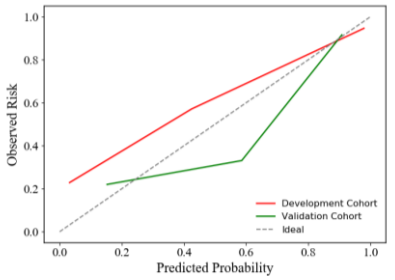


Analysis

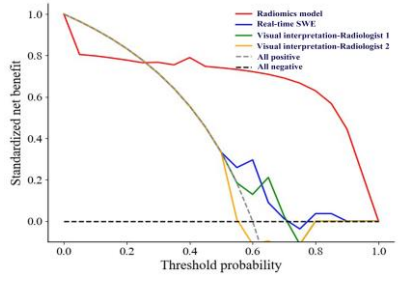
Discrimination

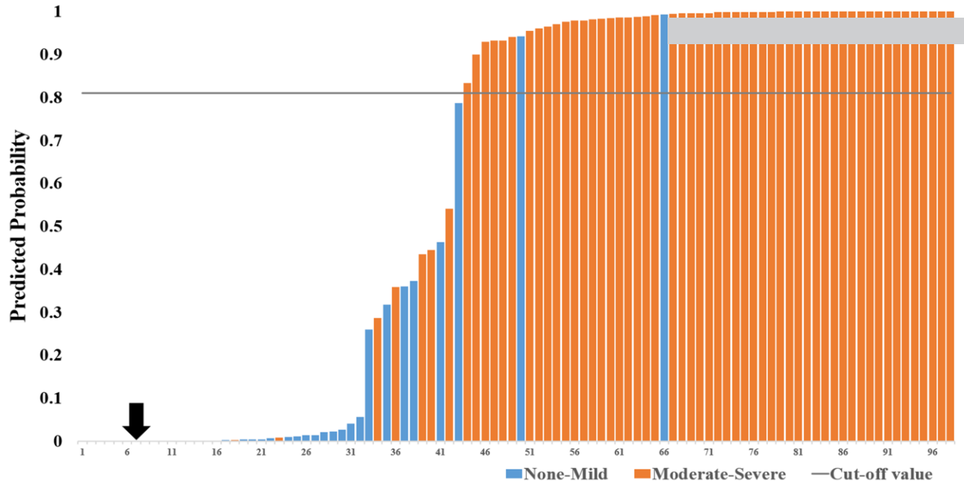


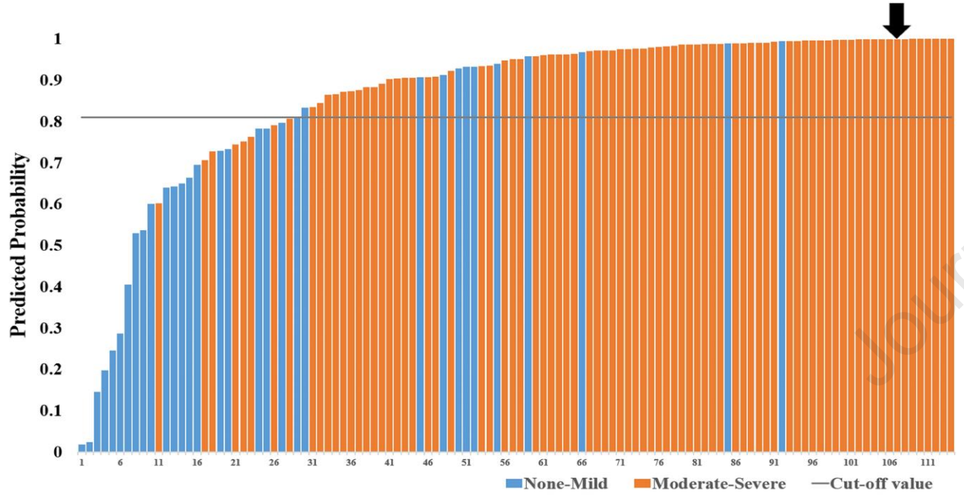
Calibration

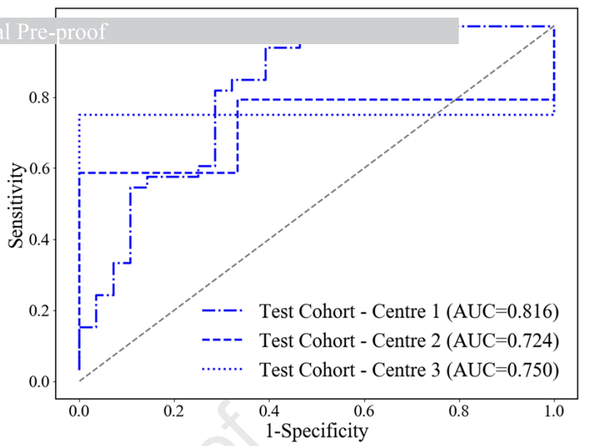


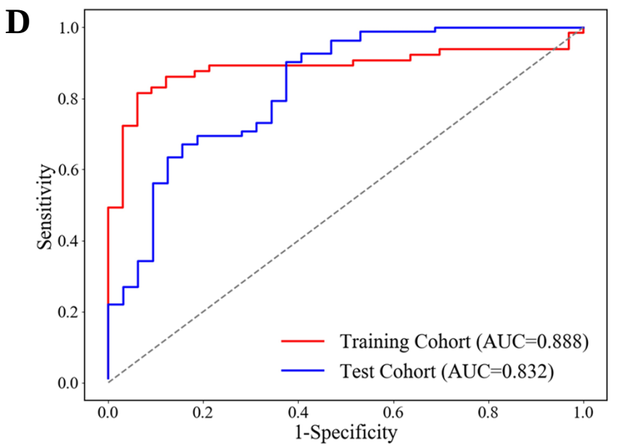
Clinical Utility

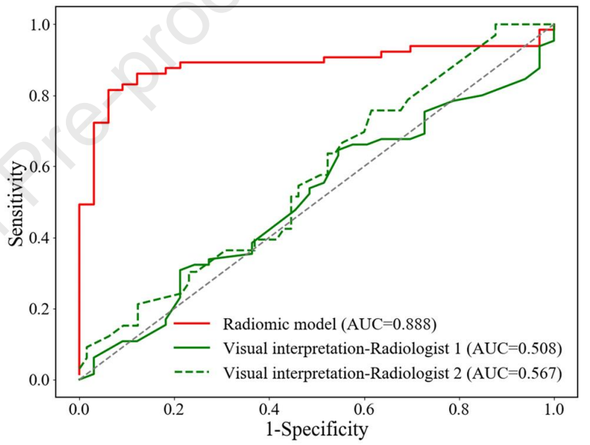


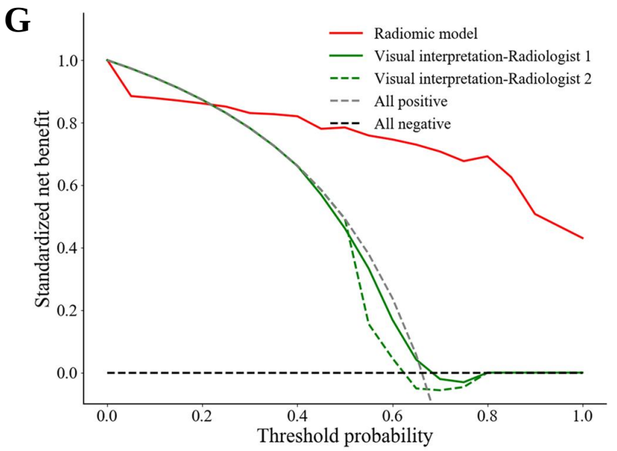


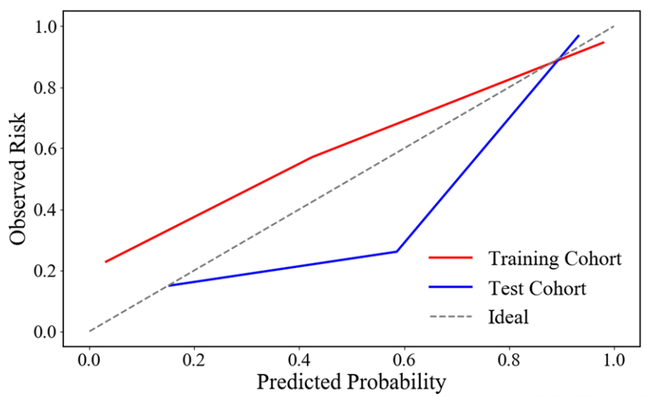


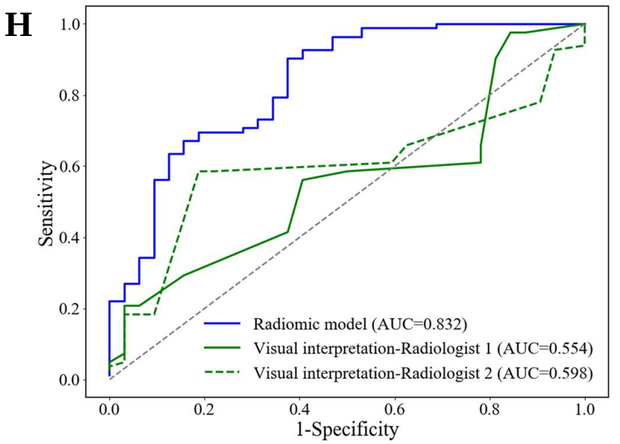


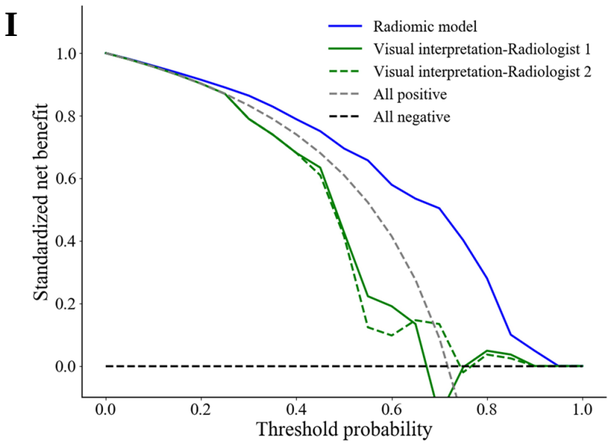












What You Need to Know

Background and Context

Accurate characterisation of intestinal fibrostenosis is crucial for the management of Crohn's disease (CD); however, no consensus currently exists regarding the most reliable method for evaluating intestinal fibrosis in CD.

New Findings

A novel computed-tomography enterography (CTE)-based radiomic model allowed for accurate characterisation of intestinal fibrosis in CD with remarkable robustness across different centres. Diagnostic performance was not affected by coexisting inflammation.

Limitations

The study is limited by its retrospective design. Furthermore, data from more centres is needed to confirm the diagnostic performance of this radiomic model.

Impact

This radiomic model improves the performance of CTE for transmural fibrosis quantitation in patients with CD; the model may be useful for improving patient care while decreasing healthcare costs.

Short summary

Our imaging model allows for accurate characterisation of intestinal fibrosis in Crohn's disease. Importantly, the model was robust across different treatment centres and was not affected by coexisting inflammation.